UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/649,909	08/26/2003	Satyanarayana Dharanipragada	N0484.70762US00	5755

23628 7590 10/25/2010

WOLF GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

EXAMINER	
COLUCCI, MICHAEL C	

ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
10/25/2010	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/649,909	DHARANIPRAGADA ET AL.	
	Examiner	Art Unit	
	MICHAEL C. COLUCCI	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *09 September 2010*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-15 and 17-27* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-15 and 17-27* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☒ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

Continued Examination Under 37 CFR 1.114

1. A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on 09/09/2010 has been entered.

Response to Arguments

2. Applicant's arguments, see Remarks, filed09/09/2010, with respect to the rejection(s) of claim(s) 1-15 and 17-27 under 35 USC 103(a) have been fully considered and are persuasive. Therefore, the rejection has been withdrawn. However, upon further consideration, a new ground(s) of rejection is made in view of Shao US 20030144841 A1. Examiner believes that though Neti *in view of* Naito teach both dependent and independent models, *creating* a model versus analyzing an existing model are two different operations. Though it may be obvious, Examiner incorporates Shao, wherein Shao teaches the creation of both speech dependent and independent models used to differentiate between male and female for example. The teachings of Shao allow for a model creation system and can easily incorporate the techniques separating female from male data taught by Neti, Naito, and Kanevsky.

Further, Examiner maintains the use of Neti in view of Wark and Naito for claims 17, 21, and 24, wherein Wark teaches the mathematical operations directed to computing accumulated confidence scores, whereby Wark improves the common and uncommon data sets of Neti. Further, having a homogenous audio segment is irrelevant, since the steps performed still classify each audio frame like the present invention. For instance a homogenous audio segment may contain diverse frames or clips, hence classification of data. Wark teaches classification of homogeneous segments, a number of statistical features are extracted from each segment. Whilst previous systems extract from each segment a feature vector, and then classify the segments based on the distribution of the feature vectors, method 200 divides each homogenous segment into a number of smaller sub-segments, or clips hereinafter, with each clip large enough to extract a meaningful feature vector f from the clip. The clip feature vectors f are then used to classify the segment from which it is extracted based on the characteristics of the distribution of the clip feature vectors f. The key advantage of extracting a number of feature vectors f from a series of smaller clips rather than a single feature vector for a whole segment is that the characteristics of the distribution of the feature vectors f over the segment of interest may be examined. Thus, whilst the signal characteristics over the length of the segment are expected to be reasonably consistent, by virtue of the segmentation algorithm, some important characteristics in the distribution of the feature vectors f over the segment of interest is significant for classification purposes (Wark [0094])

Further, Wark teaches the ability to decide whether the segment should be assigned the label of the class with the highest score, or labeled us "unknown", a confidence score is calculated. This is achieved by taking the difference of the top two model scores .sub.p and .sub.q, and normalizing that difference by the distance measure D.sub.pq between their class models p and q. This is based on the premise that an easily identifiable segment should be a lot closer to the model it belongs to than the next closest model. With further apart models, the model scores .sub.c should also be well separated before the segment is assigned the class label of the class with the highest score (Wark [0146] & Fig. 4, adjacent, previous and current segment/frame).

The substitution of class models of Wark with gender phoneme models of Neti allows for a defined class (i.e. male, female, or both).

Specification

3. The specification is objected to as failing to provide proper antecedent basis for the claimed subject matter. See 37 CFR 1.75(d)(1) and MPEP § 608.01(o). Correction of the following is required:

Claims 1-5 and 17-20 disclose a “computer readable medium” with no description within the disclosure. There is also no suggestion of a computer readable medium such as RAM, ROM, etc.

Claims 11-15 and 24-27 disclose a “computer program product” and “computer memory” with no description within the disclosure. There is also no suggestion of a

computer program product and computer memory such as program code, software stored on hardware, ROM, RAM, hard drive, etc.

Claim Rejections - 35 USC § 112

4. The following is a quotation of the first paragraph of 35 U.S.C. 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

5. Claims 1-5, 11-15, 17-20, and 24-27 rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention.

Claims 1-5 and 17-20 disclose a "computer readable medium" with no description within the disclosure. There is also no suggestion of a computer readable medium such as RAM, ROM, etc.

Claims 11-15 and 24-27 disclose a "computer program product" and "computer memory" with no description within the disclosure. There is also no suggestion of a computer program product and computer memory such as program code, software stored on hardware, ROM, RAM, hard drive, etc.

Claim Rejections - 35 USC § 103

6. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subj7ect matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

7. Claims 1-16 are rejected under 35 U.S.C. 103(a) as being unpatentable over Neti et al. US 5953701 A (hereinafter Neti) in view of Naito et al. US 5983178 A (hereinafter Naito) and further in view of Kanevsky et al. US 6529902 (hereinafter Kanevsky) and further in view of Shao US 20030144841 A1 (hereinafter Shao).

Re claims 1, 6, and 11, Neti teaches a method for generating a speech recognition model, the method comprising:

receiving female speech training data (Abstract, Col. 3 lines 37-49, Col. 4 lines 10-29, male training data, female training data, gender specific phone state models);

generating female phoneme models based on the female speech training data (Abstract, Col. 3 lines 37-49, Col. 4 lines 10-29, male training data, female training data, gender specific phone state models);

receiving male speech training data (Abstract, Col. 3 lines 37-49, Col. 4 lines 10-29, male training data, female training data, gender specific phone state models);

generating male phoneme models based on the male speech training data (Abstract, Col. 3 lines 37-49, Col. 4 lines 10-29, male training data, female training data, gender specific phone state models);

determining a difference between each female phoneme model and each corresponding male phoneme model (Abstract, Col. 3 lines 37-49, Col. 4 lines 10-29, aligning data with gender independent data, male training data, female training data, gender specific phone state models)

creating a gender-independent phoneme model

when the difference between the compared female phoneme model and the corresponding male phoneme model is less than predetermined value

However, Neti fails to teach a gender-independent/dependent phoneme model

Naito improves the model of Neti by incorporating gender impendent phonemic models such as Naito teaches a clustering processor for training a predetermined initial hidden Markov model using a predetermined training algorithm based on the speech waveform data of speakers respectively belonging to the generated K clusters, which is stored in said first storage unit, thereby generating a plurality of K hidden Markov models corresponding to the plurality of K clusters

a second storage unit for storing the plurality of K hidden Markov models generated by said clustering processor;

a first speech recognition unit for recognizing speech of an inputted uttered speech signal of a recognition-target speaker with reference to a predetermined

speaker independent phonemic hidden Markov model, and outputting a series of speech-recognized phonemes;

a speaker model selector for recognizing the speech of the inputted uttered speech signal, respectively, with reference to the plurality of K hidden Markov models stored in said second storage unit, based on the sequence of speech-recognized phonemes outputted from said first speech recognition unit, thereby calculating K likelihoods corresponding to the K hidden Markov models, and for selecting at least one hidden Markov model having the largest likelihood from the K hidden Markov models (Naito Col. 3 line 49—Col. 4 line 12).

Further, Naito teaches the recognition of phoneme dependent data which verifies whether data is independent of dependent, for example whether incoming data is within the range of a model or not (Naito Col. 15 line 54 - Col. 16 line 25).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti to incorporate a gender-independent/dependent phoneme model as taught by Naito to allow for the selection of the best combination of phoneme models with the highest probability of having correctly recognized gender based speech in a phonemic model (Naito Col. 15 line 54 - Col. 16 line 25).

However, Neti in view of Naito fails to teach creating a gender-independent phoneme model when the difference between the compared female phoneme model and the corresponding male phoneme model is less than predetermined value

adding, based on at least one criteria,

one of the gender-independent phoneme model, OR

both the female phoneme model and the corresponding male phoneme model to the speech recognition model

Kanevsky can easily substitute male and female for topics when executing a Kullback-Liebler distance method, wherein Kanevsky teaches referring to FIG. 5, which illustrates on one-way direction process of separating features belonging to different topics and topic identification via a Kullback-Liebler distance method, texts that are labeled with different topics are denoted as 501 (e.g., topic 1), 502 (e.g., topic 2), 503 (e.g., topic 3), 504 (e.g., topic N) etc. Textual features can be represented as frequencies of words, a combination of two words, a combination of three words etc. On these features, one can define metrics that allow computation of a distance between different features. For example, if topics T.sub.i give rise to probabilities P(w.sub.t.vertline.T.sub.t), where w.sub.t run all words in some vocabulary, then a distance between two topics T.sub.i and T.sub.j can be computed as #EQU13##. Using Kullback-Liebler distances is consistent with likelihood ratio criteria that are considered above, for example, in Equation (6). Similar metrics could be introduced on tokens that include T-gram words or combination of tokens, as described above. Other features reflecting topics (e.g., key words) can also be used. For every subset of k features, one can define a k dimensional vector. Then, for two different k sets, one can define a Kullback-Liebler distance using frequencies of these k sets. Using Kullback-Liebler distance, one can check which pairs of topics are sufficiently separated from each other.

Topics that are close in this metric could be combined together. For example, one can find that topics related to "LOAN" and "BANKS" are close in this metric, and therefore should be combined under a new label (e.g. "FINANCE"). Also, using these metrics, one can identify in each topic domain textual feature vectors ("balls") that are sufficiently separated from other "balls" in topic domains. These "balls" are shown in FIG. 5 as 505, 506, 504, etc. When such "balls" are identified, likelihood ratios as in FIG. 1, are computed for tokens from these "balls". (Kanevsky Col. 12 lines 15-56)

Further, Kanevsky teaches another instance of detecting whether a threshold is breached and topic similarity based on training data (Kanevsky Col. 13 lines 7-12 & lines 42-45).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti in view of Naito to incorporate a gender-independent phoneme model when the difference between the compared female phoneme model and the corresponding male phoneme model is less than predetermined value and adding, based on at least one criteria, one of the gender-independent phoneme model, or both the female phoneme model and the corresponding male phoneme model to the speech recognition model as taught by Kanevsky to allow for the generation of combined data models with similar context such as male and female together (e.g. LOAN and BANKS) and also isolated data such as explicit male and female data (e.g. medical and legal), wherein topics are labeled as a group of phonemes or unigrams utilizing a Kullback-Liebler distance, where one can check which pairs of topics are sufficiently separated from each other provided a subset

of k features, that one can define a k dimensional vector allowing computation of a distance between different features in the form of a trained group of model (Kanevsky Col. 12 lines 15-56).

However, Neti in view of Naito and Kanevsky fails to teach creating a gender-independent phoneme model

Shao teaches that speech models generated by the model generation module 25 could be of any conventional form. Specifically, it will be appreciated that different types of speech model could be created for example continuous or discreet speech models could be created. Similarly, speaker independent or speaker dependent speech models could be created. Further, the speech models themselves could be generated using any conventional algorithms. Alternatively, a number of different speech models could be created from the same selected set of utterances so that the effectiveness of different algorithms could be assessed (Shao [0136] & Fig. 2 model generation 25, model database).

Further, Shao teaches that results of testing generated speech models could also be of different forms. Instead of merely identifying correct and incorrect recognitions, confidence scores or the like for recognitions could be included in a testing report. Alternatively instead of matching an utterance with only a single word, a set of top matches could be indicated with the closeness of match for each utterance being indicated so that the amount of confusion between different words could be assessed (Shao [0137]).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti in view of Naito and Kanevsky to incorporate creating a gender-independent phoneme model as taught by Shao to allow for a model generation module based on other models (Shao Fig. 2), wherein confidence scores or the like for recognitions could be included in a testing report having a set of top matches that could be indicated with the closeness of match for each utterance being indicated so that the amount of confusion between different words could be assessed (Shao [0137]), thereby selecting the best models for creating an independent model such as a gender-independent phoneme model of Neti.

Re claims 2, 7, and 12, Naito fails to teach the method at least one computer readable medium of claim 1, wherein the at least one criteria comprises a threshold value or an upper limit for the total number of phoneme models in the speech recognition model.

Naito improves the model of Neti by incorporating gender impendent phonemic models such as Naito teaches a clustering processor for training a predetermined initial hidden Markov model using a predetermined training algorithm based on the speech waveform data of speakers respectively belonging to the generated K clusters, which is stored in said first storage unit, thereby generating a plurality of K hidden Markov models corresponding to the plurality of K clusters

a second storage unit for storing the plurality of K hidden Markov models generated by said clustering processor;

a first speech recognition unit for recognizing speech of an inputted uttered speech signal of a recognition-target speaker with reference to a predetermined speaker independent phonemic hidden Markov model, and outputting a series of speech-recognized phonemes;

a speaker model selector for recognizing the speech of the inputted uttered speech signal, respectively, with reference to the plurality of K hidden Markov models stored in said second storage unit, based on the sequence of speech-recognized phonemes outputted from said first speech recognition unit, thereby calculating K likelihoods corresponding to the K hidden Markov models, and for selecting at least one hidden Markov model having the largest likelihood from the K hidden Markov models (Naito Col. 3 line 49—Col. 4 line 12).

Further, Naito teaches the recognition of phoneme dependent data which verifies whether data is independent of dependent, for example whether incoming data is within the range of a model or not (Naito Col. 15 line 54 - Col. 16 line 25).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti to incorporate a plurality of phoneme models as taught by Naito to allow for the selection of the best combination of phoneme models with the highest probability of having correctly recognized gender based speech in a phonemic model (Naito Col. 15 line 54 - Col. 16 line 25).

However, Neti in view of Naito fails to teach a threshold value or an upper limit for the total number of phoneme models in the speech recognition model

Kanevsky teaches referring to FIG. 5, which illustrates on one-way direction process of separating features belonging to different topics and topic identification via a Kullback-Liebler distance method, texts that are labeled with different topics are denoted as 501 (e.g., topic 1), 502 (e.g., topic 2), 503 (e.g., topic 3), 504 (e.g., topic N) etc. Textual features can be represented as frequencies of words, a combination of two words, a combination of three words etc. On these features, one can define metrics that allow computation of a distance between different features. For example, if topics T.sub.i give rise to probabilities P(w.sub.t.vertline.T.sub.t), where w.sub.t run all words in some vocabulary, then a distance between two topics T.sub.i and T.sub.j can be computed as #EQU13##. Using Kullback-Liebler distances is consistent with likelihood ratio criteria that are considered above, for example, in Equation (6). Similar metrics could be introduced on tokens that include T-gram words or combination of tokens, as described above. Other features reflecting topics (e.g., key words) can also be used. For every subset of k features, one can define a k dimensional vector. Then, for two different k sets, one can define a Kullback-Liebler distance using frequencies of these k sets. Using Kullback-Liebler distance, one can check which pairs of topics are sufficiently separated from each other. Topics that are close in this metric could be combined together. For example, one can find that topics related to "LOAN" and "BANKS" are close in this metric, and therefore should be combined under a new label (e.g. "FINANCE"). Also, using these metrics, one can identify in each topic domain

textual feature vectors ("balls") that are sufficiently separated from other "balls" in topic domains. These "balls" are shown in FIG. 5 as 505, 506, 504, etc. When such "balls" are identified, likelihood ratios as in FIG. 1, are computed for tokens from these "balls". (Kanevsky Col. 12 lines 15-56)

Further, Kanevsky teaches another instance of detecting whether a threshold is breached and topic similarity based on training data (Kanevsky Col. 13 lines 7-12 & lines 42-45).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti in view of Naito to incorporate a threshold value or an upper limit for the total number of phoneme models in the speech recognition model as taught by Kanevsky to allow for the generation of combined data models with similar context such as male and female together (e.g. LOAN and BANKS) and also isolated data such as explicit male and female data (e.g. medical and legal), wherein topics are labeled as a group of phonemes or unigrams utilizing a Kullback-Liebler distance, where one can check which pairs of topics are sufficiently separated from each other provided a subset of k features, that one can define a k dimensional vector allowing computation of a distance between different features in the form of a trained group of model (Kanevsky Col. 12 lines 15-56 Fig. 5 topic clusters).

Re claims 3, 8, and 13, Neti fails to teach the method of claim 1, wherein determining the difference includes calculating a Kullback Leibler distance between the each female phoneme model and the each corresponding male phoneme model.

However, Neti fails to teach a gender-independent/dependent phoneme model

Naito improves the model of Neti by incorporating gender impendent phonemic models such as Naito teaches a clustering processor for training a predetermined initial hidden Markov model using a predetermined training algorithm based on the speech waveform data of speakers respectively belonging to the generated K clusters, which is stored in said first storage unit, thereby generating a plurality of K hidden Markov models corresponding to the plurality of K clusters

a second storage unit for storing the plurality of K hidden Markov models generated by said clustering processor;

a first speech recognition unit for recognizing speech of an inputted uttered speech signal of a recognition-target speaker with reference to a predetermined speaker independent phonemic hidden Markov model, and outputting a series of speech-recognized phonemes;

a speaker model selector for recognizing the speech of the inputted uttered speech signal, respectively, with reference to the plurality of K hidden Markov models stored in said second storage unit, based on the sequence of speech-recognized phonemes outputted from said first speech recognition unit, thereby calculating K likelihoods corresponding to the K hidden Markov models, and for selecting at least one

hidden Markov model having the largest likelihood from the K hidden Markov models (Naito Col. 3 line 49—Col. 4 line 12).

Further, Naito teaches the recognition of phoneme dependent data which verifies whether data is independent of dependent, for example whether incoming data is within the range of a model or not (Naito Col. 15 line 54 - Col. 16 line 25).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti to incorporate a gender-independent/dependent phoneme model as taught by Naito to allow for the selection of the best combination of phoneme models with the highest probability of having correctly recognized gender based speech in a phonemic model (Naito Col. 15 line 54 - Col. 16 line 25).

However, Neti in view of Naito fails to teach determining the difference includes calculating a Kullback Leibler distance

Kanevsky teaches referring to FIG. 5, which illustrates on one-way direction process of separating features belonging to different topics and topic identification via a Kullback-Liebler distance method, texts that are labeled with different topics are denoted as 501 (e.g., topic 1), 502 (e.g., topic 2), 503 (e.g., topic 3), 504 (e.g., topic N) etc. Textual features can be represented as frequencies of words, a combination of two words, a combination of three words etc. On these features, one can define metrics that allow computation of a distance between different features. For example, if topics T.sub.i give rise to probabilities P(w.sub.t.vertline.T.sub.t), where w.sub.t run all words

in some vocabulary, then a distance between two topics T.sub.i and T.sub.j can be computed as #EQU13##. Using Kullback-Liebler distances is consistent with likelihood ratio criteria that are considered above, for example, in Equation (6). Similar metrics could be introduced on tokens that include T-gram words or combination of tokens, as described above. Other features reflecting topics (e.g., key words) can also be used. For every subset of k features, one can define a k dimensional vector. Then, for two different k sets, one can define a Kullback-Liebler distance using frequencies of these k sets. Using Kullback-Liebler distance, one can check which pairs of topics are sufficiently separated from each other. Topics that are close in this metric could be combined together. For example, one can find that topics related to "LOAN" and "BANKS" are close in this metric, and therefore should be combined under a new label (e.g. "FINANCE"). Also, using these metrics, one can identify in each topic domain textual feature vectors ("balls") that are sufficiently separated from other "balls" in topic domains. These "balls" are shown in FIG. 5 as 505, 506, 504, etc. When such "balls" are identified, likelihood ratios as in FIG. 1, are computed for tokens from these "balls". (Kanevsky Col. 12 lines 15-56)

Further, Kanevsky teaches another instance of detecting whether a threshold is breached and topic similarity based on training data (Kanevsky Col. 13 lines 7-12 & lines 42-45).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti in view of Naito to incorporate determining the difference includes calculating a Kullback Leibler distance as taught by

Kanevsky to allow for the generation of combined data models with similar context such as male and female together (e.g. LOAN and BANKS) and also isolated data such as explicit male and female data (e.g. medical and legal), wherein topics are labeled as a group of phonemes or unigrams utilizing a Kullback-Liebler distance, where one can check which pairs of topics are sufficiently separated from each other provided a subset of k features, that one can define a k dimensional vector allowing computation of a distance between different features in the form of a trained group of model (Kanevsky Col. 12 lines 15-56).

However, Neti in view of Naito and Kanevsky fails to teach creating a gender-independent phoneme model

Shao teaches that speech models generated by the model generation module 25 could be of any conventional form. Specifically, it will be appreciated that different types of speech model could be created for example continuous or discreet speech models could be created. Similarly, speaker independent or speaker dependent speech models could be created. Further, the speech models themselves could be generated using any conventional algorithms. Alternatively, a number of different speech models could be created from the same selected set of utterances so that the effectiveness of different algorithms could be assessed (Shao [0136] & Fig. 2 model generation 25, model database).

Further, Shao teaches that results of testing generated speech models could also be of different forms. Instead of merely identifying correct and incorrect recognitions,

confidence scores or the like for recognitions could be included in a testing report. Alternatively instead of matching an utterance with only a single word, a set of top matches could be indicated with the closeness of match for each utterance being indicated so that the amount of confusion between different words could be assessed (Shao [0137]).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti in view of Naito and Kanevsky to incorporate creating a gender-independent phoneme model as taught by Shao to allow for a model generation module based on other models (Shao Fig. 2), wherein confidence scores or the like for recognitions could be included in a testing report having a set of top matches that could be indicated with the closeness of match for each utterance being indicated so that the amount of confusion between different words could be assessed (Shao [0137]), thereby selecting the best models for creating an independent model such as a gender-independent phoneme model of Neti.

Re claims 4, 9, and 14, Neti in view of Naito fails to teach the method of claim 3, wherein the difference is a threshold Kullback Leibler distance quantity.

Kanevsky teaches the Kullback-Leibler distance (Kanevsky Col. 5, lines 9-11) between any two topics is at least h, where h ~s some sufficiently large threshold, also Kanevsky teaches (Kanevsky Col. 12, lines 44-47) that while using the Kullback-Leibler

distance, one can check which pairs of topics are sufficiently separated from each other, and that topics that are close in this metric could be combined together).

Kanevsky also explicitly teaches how a difference is sufficient, such as classifying data groups when compared, and also creating independence from classificaiton if there is no topic discovered (Kanevsky Col. 5 lines 8-25).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti in view of Naito to incorporate whether the model information is insignificant is based on a threshold Kullback Leibler distance quantity as taught by Kanevsky to allow for an improved language modeling for automatic speech decoding and differentiation between data groups, wherein a sufficiently large threshold indicates either separate or combinational probabilities (Kanevsky Col. 2, lines 50-52).

Re claims 5, 10, and 15, Neti teaches method of claim 1, wherein the female phoneme models, male phoneme models, and gender-independent phoneme models are Gaussian mixture models (Neti Col. 3 lines 50-67).

However, Neti fails to teach a gender-independent/dependent phoneme model

Naito improves the model of Neti by incorporating gender impendent phonemic models such as Naito teaches a clustering processor for training a predetermined initial hidden Markov model using a predetermined training algorithm based on the speech waveform data of speakers respectively belonging to the generated K clusters, which is

stored in said first storage unit, thereby generating a plurality of K hidden Markov models corresponding to the plurality of K clusters

a second storage unit for storing the plurality of K hidden Markov models generated by said clustering processor;

a first speech recognition unit for recognizing speech of an inputted uttered speech signal of a recognition-target speaker with reference to a predetermined speaker independent phonemic hidden Markov model, and outputting a series of speech-recognized phonemes;

a speaker model selector for recognizing the speech of the inputted uttered speech signal, respectively, with reference to the plurality of K hidden Markov models stored in said second storage unit, based on the sequence of speech-recognized phonemes outputted from said first speech recognition unit, thereby calculating K likelihoods corresponding to the K hidden Markov models, and for selecting at least one hidden Markov model having the largest likelihood from the K hidden Markov models (Naito Col. 3 line 49—Col. 4 line 12).

Further, Naito teaches the recognition of phoneme dependent data which verifies whether data is independent of dependent, for example whether incoming data is within the range of a model or not (Naito Col. 15 line 54 - Col. 16 line 25).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti to incorporate a gender-independent/dependent phoneme model as taught by Naito to allow for the selection of the best combination of phoneme models with the highest probability of having correctly

recognized gender based speech in a phonemic model (Naito Col. 15 line 54 - Col. 16 line 25).

However, Neti in view of Naito and Kanevsky fails to teach creating a gender-independent phoneme model

Shao teaches that speech models generated by the model generation module 25 could be of any conventional form. Specifically, it will be appreciated that different types of speech model could be created for example continuous or discreet speech models could be created. Similarly, speaker independent or speaker dependent speech models could be created. Further, the speech models themselves could be generated using any conventional algorithms. Alternatively, a number of different speech models could be created from the same selected set of utterances so that the effectiveness of different algorithms could be assessed (Shao [0136] & Fig. 2 model generation 25, model database).

Further, Shao teaches that results of testing generated speech models could also be of different forms. Instead of merely identifying correct and incorrect recognitions, confidence scores or the like for recognitions could be included in a testing report. Alternatively instead of matching an utterance with only a single word, a set of top matches could be indicated with the closeness of match for each utterance being indicated so that the amount of confusion between different words could be assessed (Shao [0137]).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti in view of Naito and Kanevsky to incorporate creating a gender-independent phoneme model as taught by Shao to allow for a model generation module based on other models (Shao Fig. 2), wherein confidence scores or the like for recognitions could be included in a testing report having a set of top matches that could be indicated with the closeness of match for each utterance being indicated so that the amount of confusion between different words could be assessed (Shao [0137]), thereby selecting the best models for creating an independent model such as a gender-independent phoneme model of Neti.

8. **Claims 17-27 are rejected under 35 U.S.C. 103(a) as being unpatentable over Neti et al. US 5953701 A (hereinafter Neti) in view of Wark US 20030231775 (hereinafter Wark) and further in view of Naito et al. US 5983178 A (hereinafter Naito).**

Re claims 17, 21, and 24, Neti teaches a system for recognizing speech data from an audio stream originating from one of a plurality of data classes ([0094]), each data class having a class-dependent phoneme model, the system comprising:

a computer processor (Col. 6 lines 24-49);

a receiving module configured to receive a current feature vector of the audio stream (Col. 6 lines 24-49);

a first computing module configured to compute a current best estimates (Col. 3 lines 50-67) that the current feature vector belongs to one of the plurality of data classes (Col. 5 lines 9-21);

However, Neti fails to teach a second computing module configured to compute an accumulated confidence values for each of the plurality of data class that the current feature vector belongs to each one of the plurality of data classes, the confidence value for each data class of the plurality of data classes based on the current best estimatefor the data class and on previous confidence values for the data class, the previous confidence values associated with previous feature vectors of the audio stream;

a weighing module configured to weigh the class-dependent phoneme models based on the accumulated confidence values; and

a recognizing module configured to recognize the current feature vector (based on the weighted class-dependent phoneme models; and

Wark teaches classification of homogeneous segments, a number of statistical features are extracted from each segment. Whilst previous systems extract from each segment a feature vector, and then classify the segments based on the distribution of the feature vectors, method 200 divides each homogenous segment into a number of smaller sub-segments, or clips hereinafter, with each clip large enough to extract a meaningful feature vector f from the clip. The clip feature vectors f are then used to classify the segment from which it is extracted based on the characteristics of the distribution of the clip feature vectors f. The key advantage of extracting a number of

feature vectors f from a series of smaller clips rather than a single feature vector for a whole segment is that the characteristics of the distribution of the feature vectors f over the segment of interest may be examined. Thus, whilst the signal characteristics over the length of the segment are expected to be reasonably consistent, by virtue of the segmentation algorithm, some important characteristics in the distribution of the feature vectors f over the segment of interest is significant for classification purposes (Wark [0094])

Further, Wark teaches the ability to decide whether the segment should be assigned the label of the class with the highest score, or labeled us "unknown", a confidence score is calculated. This is achieved by taking the difference of the top two model scores .sub.p and .sub.q, and normalizing that difference by the distance measure D.sub.pq between their class models p and q. This is based on the premise that an easily identifiable segment should be a lot closer to the model it belongs to than the next closest model. With further apart models, the model scores .sub.c should also be well separated before the segment is assigned the class label of the class with the highest score (Wark [0146] & Fig. 4, adjacent, previous and current segment/frame).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti to incorporate a second computing module configured to compute an accumulated confidence values for each of the plurality of data class that the current feature vector belongs to each one of the plurality of data classes, the confidence value for each data class of the plurality of data classes based on the current best estimatefor the data class and on previous confidence values

for the data class, the previous confidence values associated with previous feature vectors of the audio stream, a weighing module configured to weigh the class-dependent phoneme models based on the accumulated confidence values, and a recognizing module configured to recognize the current feature vector (based on the weighted class-dependent phoneme models as taught by Wark to allow for normalization of a difference by a distance, whereby an easily identifiable segment should be a lot closer to the model it belongs to than the next closest model (Wark [0146]), wherein a confidence score or score is used to better classify speech, whereby segments of feature vectors are classified, making important characteristics in adjacent, current, and previous frames in the distribution of the feature vectors more apparent (Wark [0094]), wherein the best model score is achieved (Wark [0129-0130]).

However, Neti in view of Wark fails to teach creating a class-independent/dependent phoneme models

Naito improves the model of Neti by incorporating gender impendent phonemic models such as Naito teaches a clustering processor for training a predetermined initial hidden Markov model using a predetermined training algorithm based on the speech waveform data of speakers respectively belonging to the generated K clusters, which is stored in said first storage unit, thereby generating a plurality of K hidden Markov models corresponding to the plurality of K clusters

a second storage unit for storing the plurality of K hidden Markov models generated by said clustering processor;

a first speech recognition unit for recognizing speech of an inputted uttered speech signal of a recognition-target speaker with reference to a predetermined speaker independent phonemic hidden Markov model, and outputting a series of speech-recognized phonemes;

a speaker model selector for recognizing the speech of the inputted uttered speech signal, respectively, with reference to the plurality of K hidden Markov models stored in said second storage unit, based on the sequence of speech-recognized phonemes outputted from said first speech recognition unit, thereby calculating K likelihoods corresponding to the K hidden Markov models, and for selecting at least one hidden Markov model having the largest likelihood from the K hidden Markov models (Naito Col. 3 line 49—Col. 4 line 12).

Further, Naito teaches the recognition of phoneme dependent data which verifies whether data is independent of dependent, for example whether incoming data is within the range of a model or not (Naito Col. 15 line 54 - Col. 16 line 25).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti in view of Wark to incorporate class-dependent phoneme models as taught by Naito to allow for the selection of the best combination of phoneme models with the highest probability of having correctly recognized gender based speech in a phonemic model (Naito Col. 15 line 54 - Col. 16 line 25).

Re claims 18, 22, and 25, Neti teaches the method of claim 17, wherein computing the current vector probability includes estimating a posteriori class probability for the current feature vector (Col. 2 lines 1-8))

Re claims 19, 23, and 26, Neti fails to teach the method of claim 17, wherein computing the accumulated confidence level further comprising weighing the current vector probability more than the previous vector probabilities.

Wark teaches classification of homogeneous segments, a number of statistical features are extracted from each segment. Whilst previous systems extract from each segment a feature vector, and then classify the segments based on the distribution of the feature vectors, method 200 divides each homogenous segment into a number of smaller sub-segments, or clips hereinafter, with each clip large enough to extract a meaningful feature vector f from the clip. The clip feature vectors f are then used to classify the segment from which it is extracted based on the characteristics of the distribution of the clip feature vectors f. The key advantage of extracting a number of feature vectors f from a series of smaller clips rather than a single feature vector for a whole segment is that the characteristics of the distribution of the feature vectors f over the segment of interest may be examined. Thus, whilst the signal characteristics over the length of the segment are expected to be reasonably consistent, by virtue of the segmentation algorithm, some important characteristics in the distribution of the feature

vectors f over the segment of interest is significant for classification purposes (Wark [0094])

Further, Wark teaches the ability to decide whether the segment should be assigned the label of the class with the highest score, or labeled us "unknown", a confidence score is calculated. This is achieved by taking the difference of the top two model scores .sub.p and .sub.q, and normalizing that difference by the distance measure D.sub.pq between their class models p and q. This is based on the premise that an easily identifiable segment should be a lot closer to the model it belongs to than the next closest model. With further apart models, the model scores .sub.c should also be well separated before the segment is assigned the class label of the class with the highest score (Wark [0146] & Fig. 4, adjacent, previous and current segment/frame).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti to incorporate computing the accumulated confidence level further comprising weighing the current vector probability more than the previous vector probabilities as taught by Wark to allow for normalization of a difference by a distance, whereby an easily identifiable segment should be a lot closer to the model it belongs to than the next closest model (Wark [0146]), wherein a confidence score or score is used to better classify speech, whereby segments of feature vectors are classified, making important characteristics in adjacent, current, and previous frames in the distribution of the feature vectors more apparent (Wark [0094]).

Re claims 20 and 27, Neti teaches the method of claim 17, further comprising determining if another feature vector is available for analysis (Col. 6 lines 24-49).

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to MICHAEL C. COLUCCI whose telephone number is (571)270-1847. The examiner can normally be reached on 8:30 am - 5:00 pm , Monday - Friday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571)-272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Michael C Colucci/
Examiner, Art Unit 2626
Patent Examiner
AU 2626
(571)-270-1847
Examiner FAX: (571)-270-2847
Michael.Colucci@uspto.gov